BIOINFORMATICS

cience for the 21st Century **will employ increasingly sophisticated computational tools to manage and analyze the rapidly growing volume of biological data**

Bioinformatics, which was born of the convergence of the rapidly expanding fields of computer science and biomedicine in the latter part of the 20th century, will be a major biological focus for the first part of the 21st. Simply put, our ability to produce huge amounts of biological information has outrun our ability to handle it most usefully.

Biomedical research is producing millions of encyclopedias worth of information. The Human Genome Project by itself generates floods of data; by one estimate in mid-1999, the volume is doubling every 14 months. Research into many other areas — sequencing of other organisms, gene expression, protein structure delineation, population genetics, clinical trials and so on — adds to the data deluge. A major challenge now is to find ways to organize, interpret and disseminate the information effective — and that's the role of bioinformatics.

At Los Alamos, the bioinformatics age began in 1982 with the creation of GenBank, a database now maintained by the National Institutes of Health as a national repository of genetic sequencing information. In 1986, the Laboratory established another national repository that stores and distributes HIV sequencing data. In the past three or four years, the Laboratory has undertaken similar projects for influenza, sexually transmitted bacteria and viruses, and oral microbial pathogens. It also has bioinformatic projects under way for anthrax, plague and smallpox as part of the national effort to address the threat of bioterrorism.

These specialized databases, which house sequencing and related molecular data, allow first-level analyses that help researchers understand how bacteria function, discover similarities between pathogens and find ways to target drugs most efficaciously. They are also valuable in tracking evolutionary patterns, which can provide clues to the future of, for example, a pathogen. The Laboratory has developed, and continues to develop, innovative software that enables flexible, sophisticated searches of the databases.

For the immediate future, the biomedical and computational communities are seeking ways to close a knowledge gap resulting from the unexpectedly great technological advances by training more people in computational biology.

The research community also is seeking to improve interoperability, or the way in which different databases can be accessed effectively by different researchers, and data mining, or techniques to search the mountains of information for useful descriptive or predictive patterns.

Bioinformatics in the future will aid the understanding and tracking of molecular epidemics, including those that may result from bioterrorism; the study of cellular responses to infections; the rapid sequencing of bacteria and other organisms; and the interpretation of statistical data from clinical trials. New imaging technologies will help diagnose and treat a variety of diseases.

Biomedical research, with its promise of understanding how biological systems work and thereby improving human health, will continue to leap forward. Research at Los Alamos and elsewhere will help develop the computational tools needed for researchers and the medical community make the best use of the information they are collecting.

CONTACT: John A. Webster at webster@lanl.gov or (505) 667-5543. For more "Science for the 21st Century," go to http://www.lanl.gov/orgs/pa/science21 on the World Wide Web.